AI COMMUNITY CONFERENCE

Toronto, Canada
July 12, 2024

Azure AI Search and Open AI

Peter Carson

#AICommunityConference

Microsoft

Peter Carson

- President, Extranet User Manager
- 13-time Microsoft M365 MVP
- peter@envisionit.com
- blog.petercarson.ca
- President, Toronto Microsoft 365 User Group

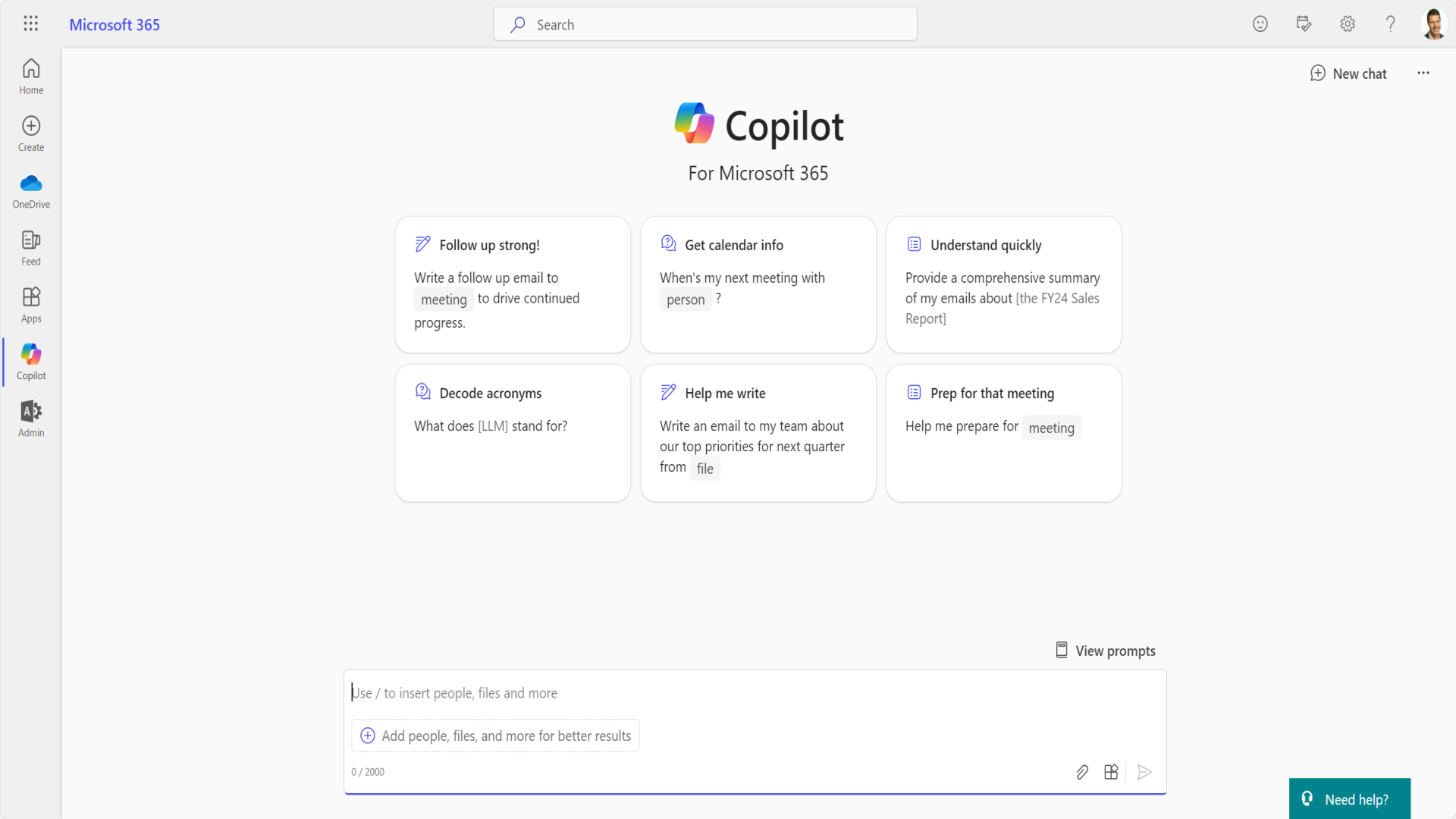
Thank You, Sponsors

PLATINUM	GRAVITY UNION SOFTEL Creospark Work how you want.
GOLD	Impro Extranet User Manager
COMMUNITY	NorthTorontoUG.com Cloud & DevOps User Group Toronto Fabric User Group MS EXCEL TORONTO Meetup AZURE & MODERN WORK User Groups Microsoft LEARN TheTechPlatform Where Tech Meets. Microsoft Reactor

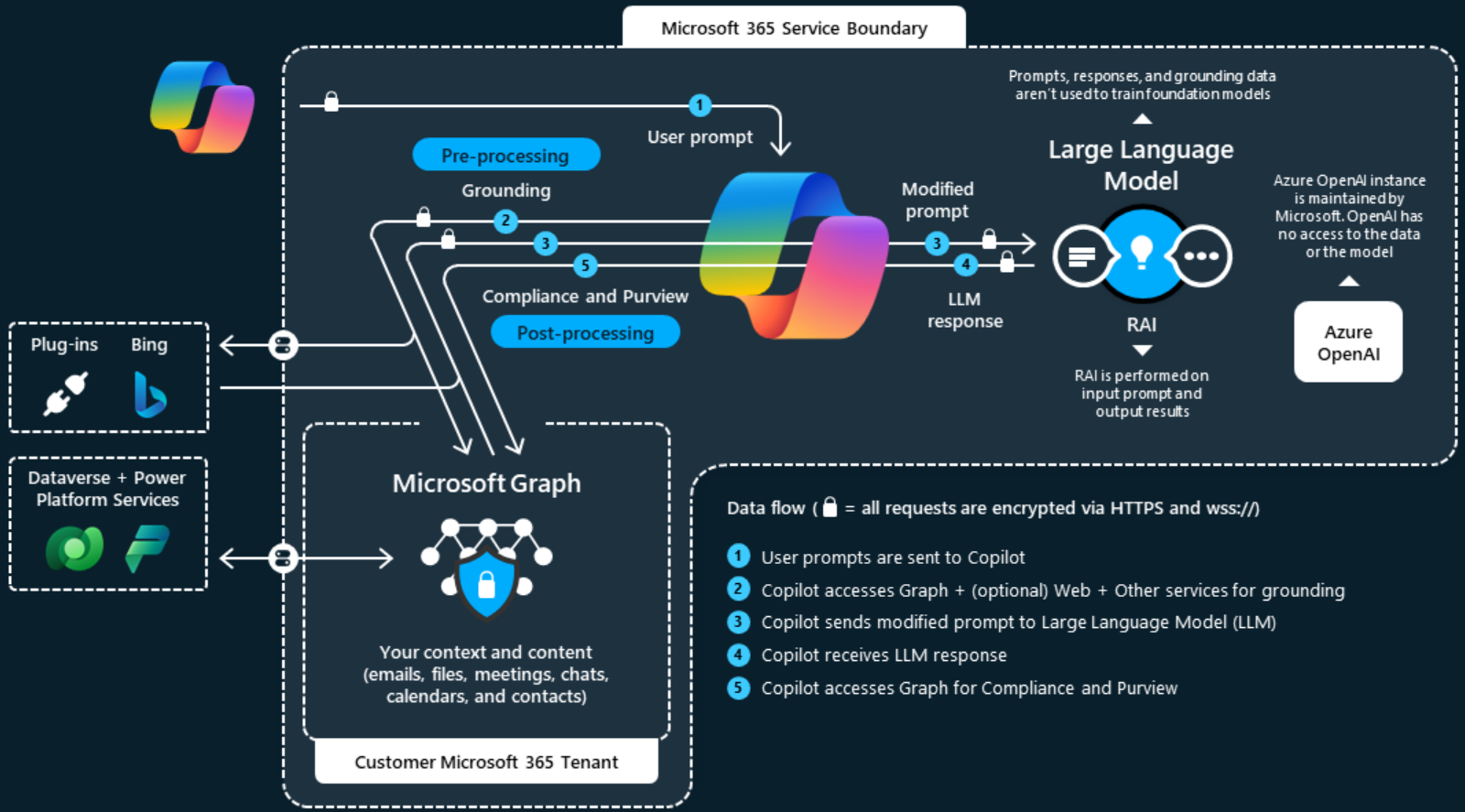
Agenda

- Copilot for Microsoft 365
- Build Your Own Chat
- Azure Open AI and AI Search
- Real World Demonstrations
- Closing and Questions

Copilot


Microsoft 365
Search
Home
Create
OneDrive
Feed
Apps
Copilot
Admin
New chat
Copilot
For Microsoft 365
Follow up strong!
Write a follow up email to meeting to drive continued progress.
Get calendar info
When's my next meeting with person ?
Understand quickly
Provide a comprehensive summary of my emails about [the FY24 Sales Report]
Decode acronyms
What does [LLM] stand for?
Help me write
Write an email to my team about our top priorities for next quarter from file
Prep for that meeting
Help me prepare for meeting
View prompts
Use / to insert people, files and more
Add people, files, and more for better results
0 / 2000
Need help?

Microsoft Copilot for Microsoft 365 architecture



Build Your Own Chat

Azure OpenAI

- Azure OpenAI Service – Large Language Models for Generative AI (microsoft.com)
- Azure OpenAI service features offers REST API to GPT-4 and GPT-35
- Both language models co-develop their APIs, guaranteeing compatibility and easy transitions
- Azure OpenAI ensures the same security standards as Microsoft Azure
- Azure OpenAI offers private networking, regional availability, and responsible AI content filtering

Chat with our support team

How to search and book rentals?

Ask now

What is the refund policy?

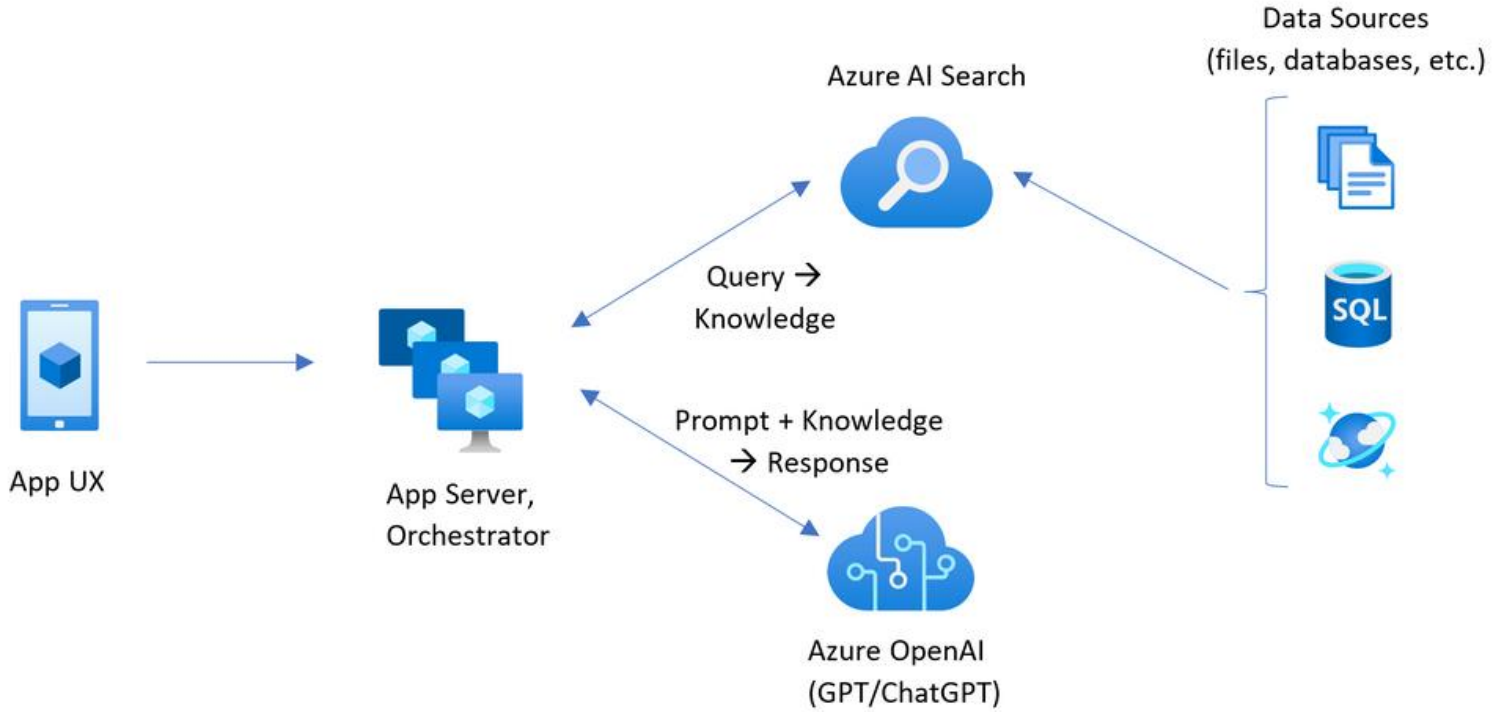
Ask now

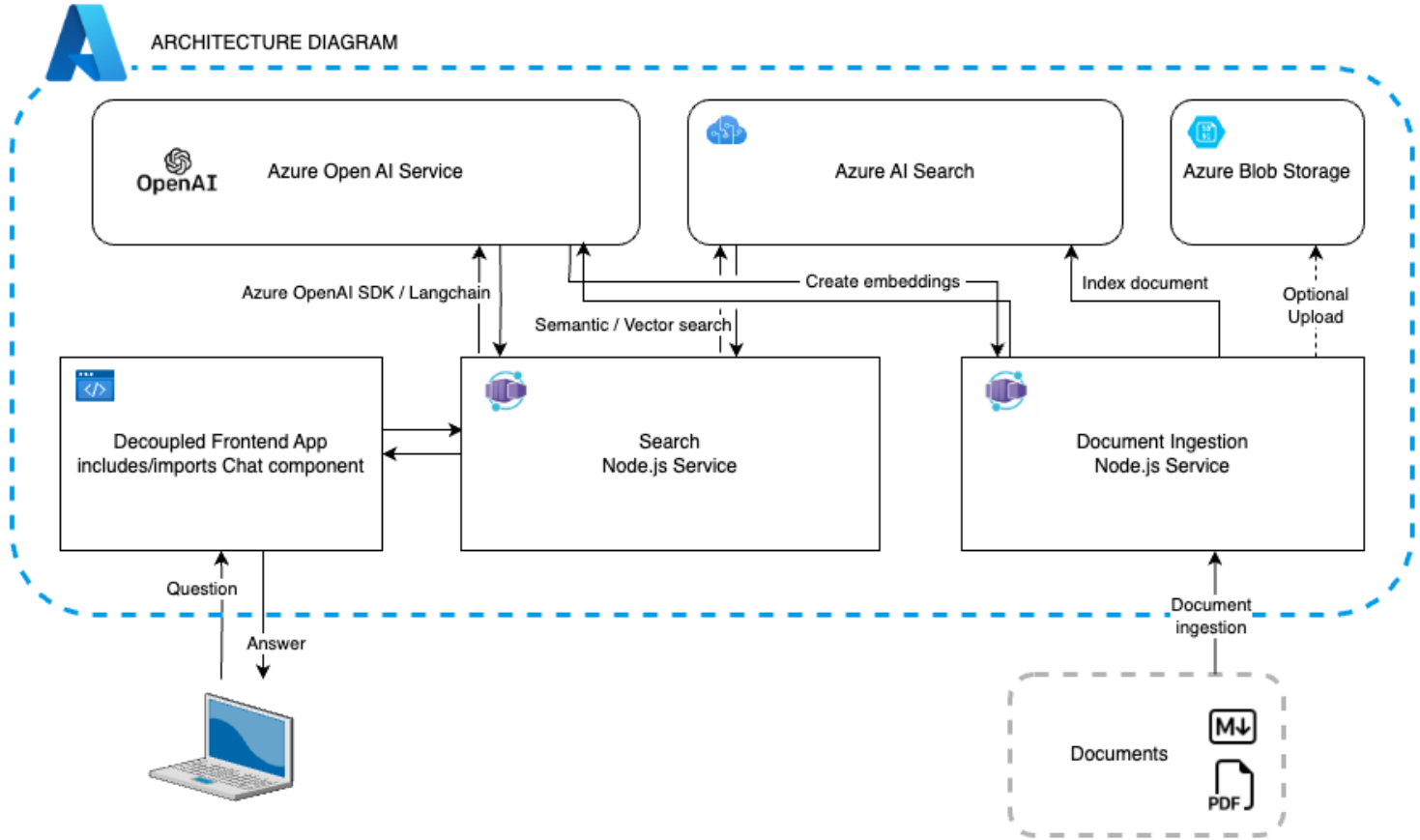
How to contact a representative?

Ask now

Type your question, eg. "How to search and book rentals?"

Architectural overview




ARCHITECTURE DIAGRAM
OpenAI
Azure Open AI Service
Azure AI Search
Azure Blob Storage
Azure OpenAI SDK / Langchain
Create embeddings
Index document
Optional Upload
Semantic / Vector search
Decoupled Frontend App
includes/imports Chat component
Search
Node.js Service
Document Ingestion
Node.js Service
Question
Answer
Document ingestion
Documents
M↓
PDF

Pricing

Azure OpenAI Service - Pricing | Microsoft Azure

Priced and scaled by tokens and context

- 1 token ~= 4 chars in English
- 1 token ~= ¾ words
- 100 tokens ~= 75 words

Models	Context	Input (Per 1,000 tokens)	Output (Per 1,000 tokens)
GPT-3.5-Turbo-0125	16K	$0.0007	$0.0021
GPT-3.5-Turbo-Instruct	4K	$0.0021	$0.003
GPT-4	8K	$0.042	$0.083
GPT-4	32K	$0.083	$0.165

Calculating Azure OpenAI Service Usage Costs: A Comprehensive Guide | by Oleksandr Krakovetskyi | The DevRain Tech Blog | Jul, 2023 | Medium

Search

Transitioning from Keyword to Semantic and Vector Search

Keyword Search: Direct matching of query terms to terms in documents or databases.

Use Cases: Early search engines, basic database queries.

Advantages: Simple, quick, easy to implement.

Limitations:

- Requires exact term matches
- Can yield irrelevant results if terms are out of context
- Can miss relevant documents if they don't use the exact terms

Semantic Search

Understands the meaning and intent behind user queries, rather than just matching keywords.

Use Cases: modern search engines, CMS, e-commerce recommendations.

Advantages:

- Relevant results
- Can answer queries instead of just returning files
- User-friendly with natural language queries

Limitations:

- Requires more sophisticated algorithms
- Potential for ambiguity in interpretation

Vector Search

Finds items in a dataset based on their similarity in meaning or context, rather than exact keyword matches.

Use Cases: advanced data retrieval systems, document clustering, and topic modelling.

Advantages:

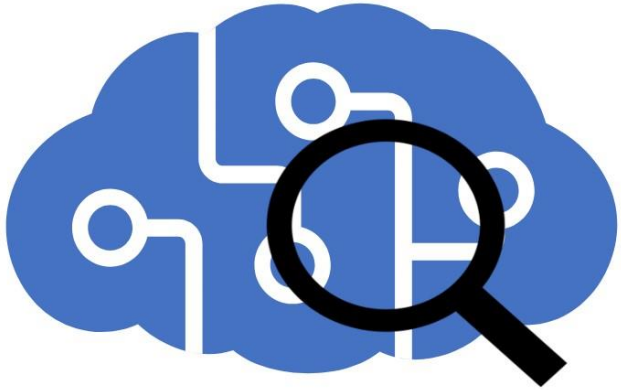
- Relevant results
- Understands context, relationships between words
- Supports languages and terms with sparse data

Limitations:

- Requires substantial computational power
- Vector models need training and refining

Azure AI Search

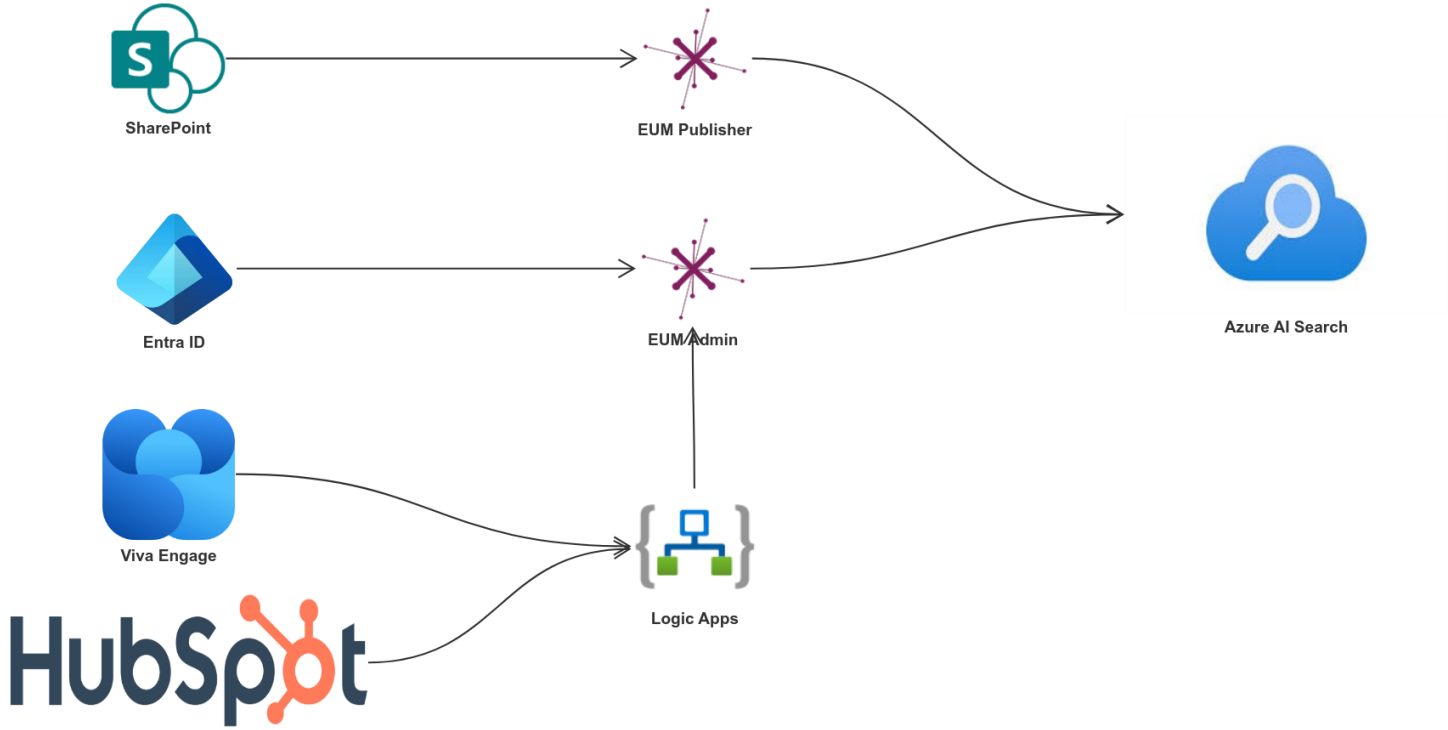
- Cloud service providing a full-text search engine, offering semantic and vector search capabilities
- Intended for developers needing application/based search experiences
- Benefits from vast data sources and AI-enhanced indexing



Azure AI Search Pricing

	Free	Basic	Standard S1	Standard S2	Standard S3	Storage Optimised L1	Storage Optimised L2
Storage[1]	50 MB	15 GB (max 45 GB per service)	160 GB (max 1.9 TB per service)	512 GB (max 6 TB per service)	1 TB (max 12 TB per service)	2 TB (max 24 TB per service)	4 TB (max 48 TB per service)
Max indexes per service	3	15	50	200	200 or 1,000/partition in high density mode	10	10
Scale out limits	N/A	Up to 9 units per service[2] (max 3 partition; max 3 replicas)	Up to 36 units per service (max 12 partition; max 12 replicas)	Up to 36 units per service (max 12 partition; max 12 replicas)	Up to 36 units per service (max 12 partition; max 12 replicas) up to 3 partitions in high density[3] mode	Up to 36 units per service (max 12 partition; max 12 replicas)	Up to 36 units per service (max 12 partition; max 12 replicas)
Price per SU (Scale Unit)	**$0**/hour	**$0.14**/hour	**$0.47**/hour	**$1.85**/hour	**$3.69**/hour	**$5.26**/hour	**$10.52**/hour

EUM AI Search Architecture



Azure AI Search Demo

ExtranetUserManager

Platform Solutions Resources Pricing About Support Test Ride PC

Home > Resources > Articles

Articles

Enter a search term

Best Practices in Secure File Sharing for Government

Jun 27, 2024

In 2023, private data exposure incidents reached a record high, with 100 reported cases impacting a massive 15 million people. With government agencies handling vast amounts of sensitive information, safeguarding data from unauthorized access has become critical.

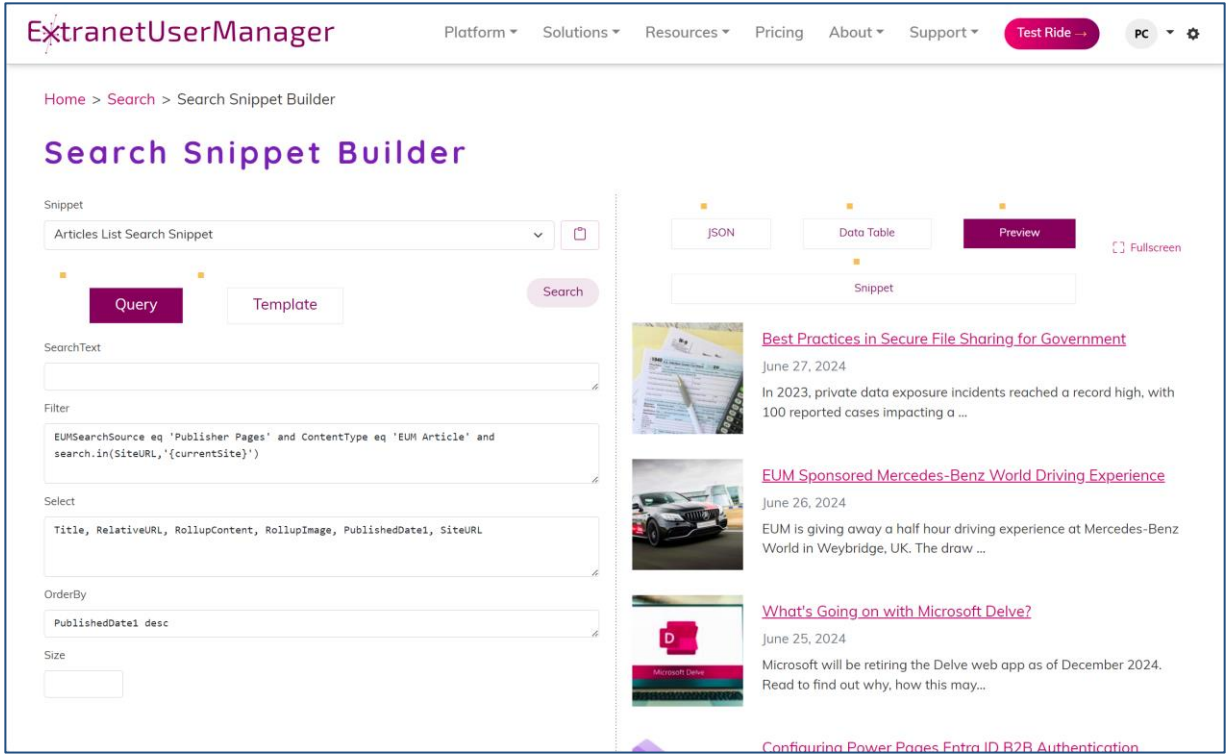
EUM Sponsored Mercedes-Benz World Driving Experience

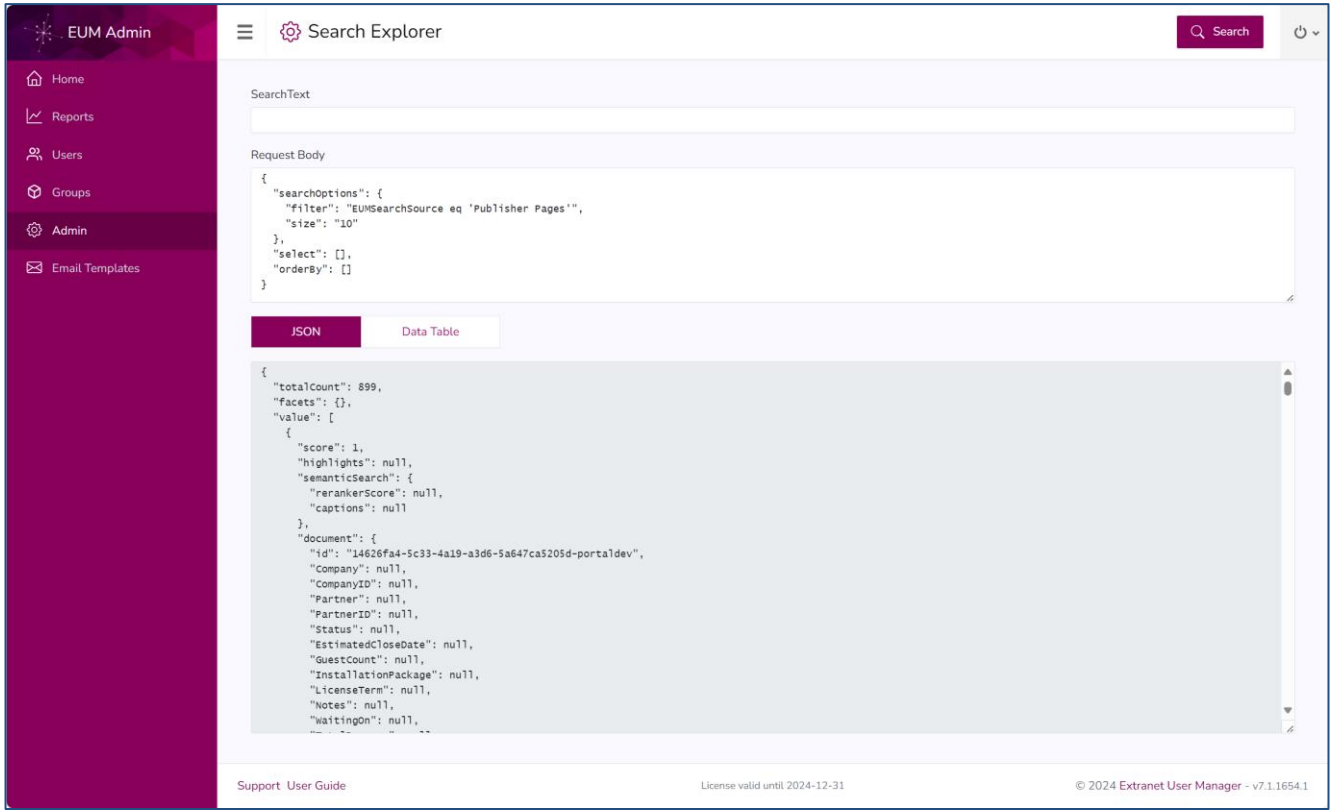
Jun 26, 2024

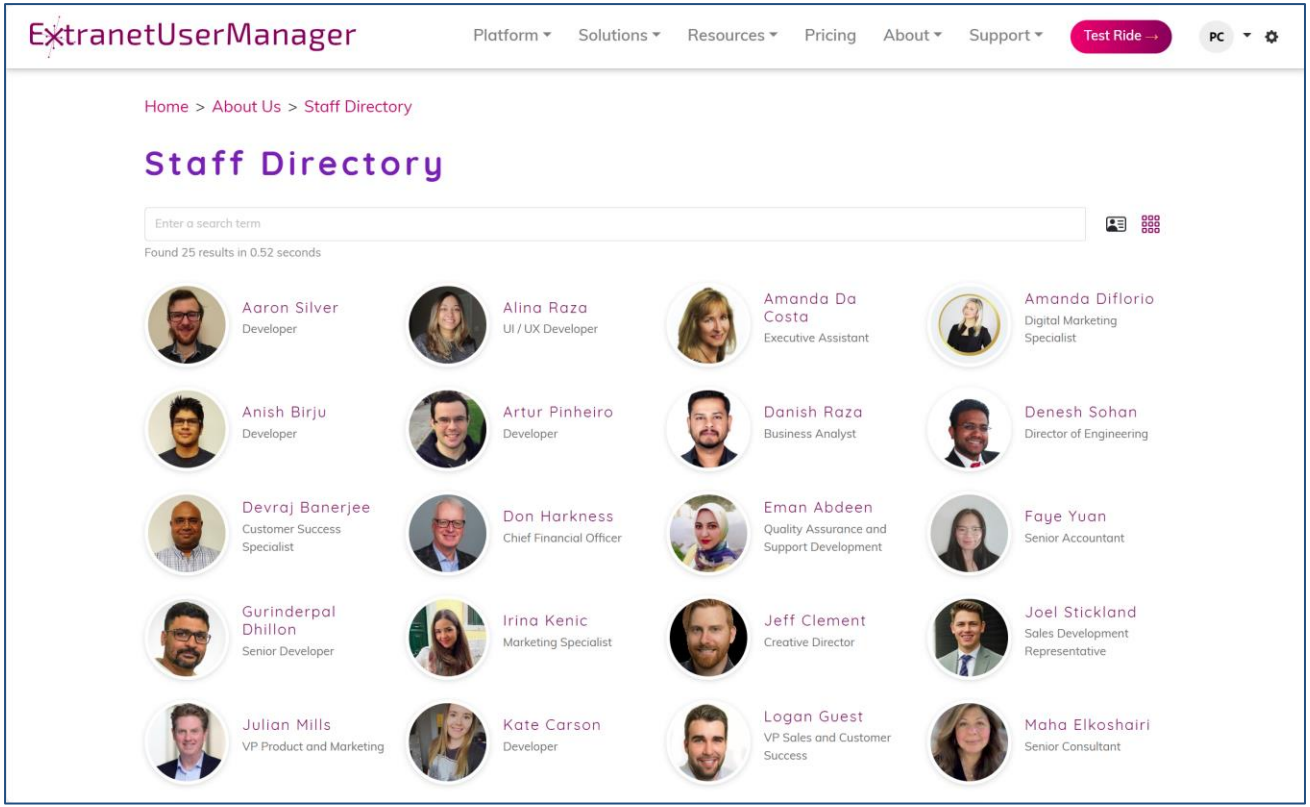
EUM is giving away a half hour driving experience at Mercedes-Benz World in Weybridge, UK. The draw will take place at the end of Peter's Teams and Microsoft 365 Sharing - Unstructured and Structured session at 12:15 PM on June 27, 2024.

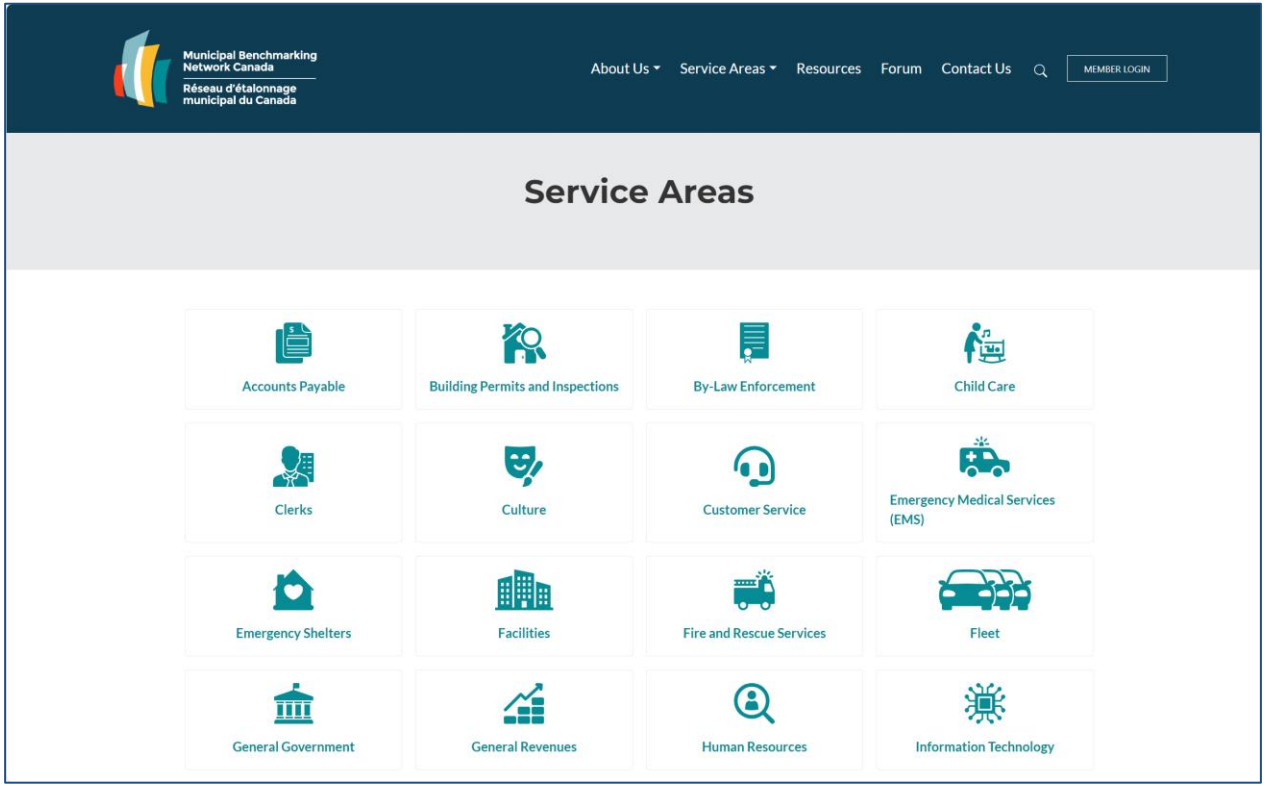
What's Going on with Microsoft Delve?

Jun 25, 2024

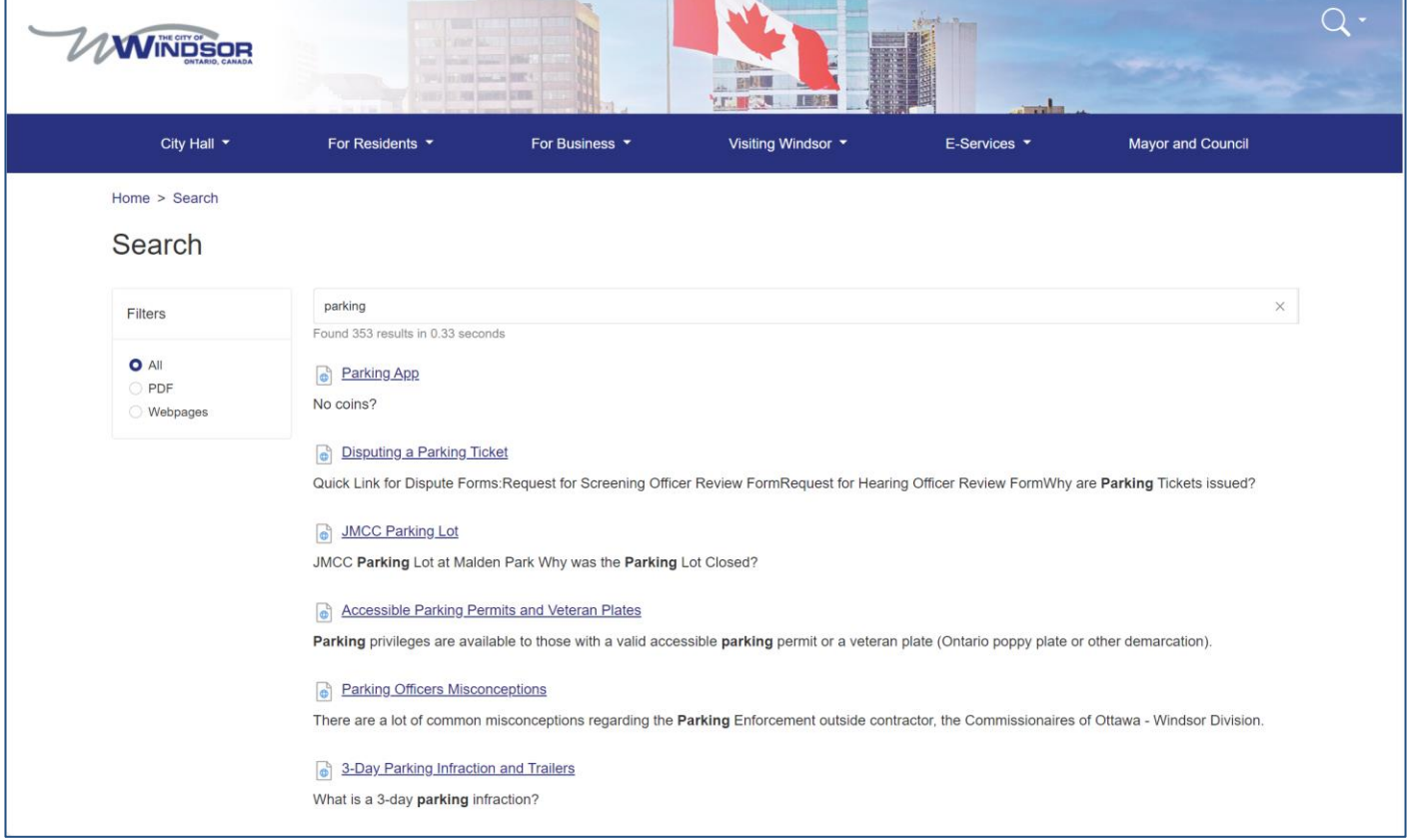

ExtranetUserManager
Platform
Solutions
Resources
Pricing
About
Support
Test Ride →
PC
Home > Search > Search Snippet Builder
Search Snippet Builder
Snippet
Articles List Search Snippet
Query
Template
Search
SearchText
Filter
EUMSearchSource eq 'Publisher Pages' and ContentType eq 'EUM Article' and search.in(SiteURL,'{currentSite}')
Select
Title, RelativeURL, RollupContent, RollupImage, PublishedDate1, SiteURL
OrderBy
PublishedDate1 desc
Size
JSON
Data Table
Preview
Fullscreen
Snippet
Best Practices in Secure File Sharing for Government
June 27, 2024
In 2023, private data exposure incidents reached a record high, with 100 reported cases impacting a ...
EUM Sponsored Mercedes-Benz World Driving Experience
June 26, 2024
EUM is giving away a half hour driving experience at Mercedes-Benz World in Weybridge, UK. The draw ...
What's Going on with Microsoft Delve?
June 25, 2024
Microsoft will be retiring the Delve web app as of December 2024. Read to find out why, how this may...


EUM Admin
Home
Reports
Users
Groups
Admin
Email Templates
Search Explorer
Search
SearchText
Request Body
{
"searchOptions": {
"filter": "EUMSearchSource eq 'Publisher Pages'",
"size": "10"
},
"select": [],
"orderBy": []
}
JSON
Data Table
{
"totalCount": 899,
"facets": {},
"value": [
{
"score": 1,
"highlights": null,
"semanticSearch": {
"rerankerScore": null,
"captions": null
},
"document": {
"id": "14626fa4-5c33-4a19-a3d6-5a647ca5205d-portaldev",
"Company": null,
"CompanyID": null,
"Partner": null,
"PartnerID": null,
"Status": null,
"EstimatedCloseDate": null,
"GuestCount": null,
"InstallationPackage": null,
"LicenseTerm": null,
"Notes": null,
"WaitingOn": null,
Support User Guide
License valid until 2024-12-31
© 2024 Extranet User Manager - v7.1.1654.1

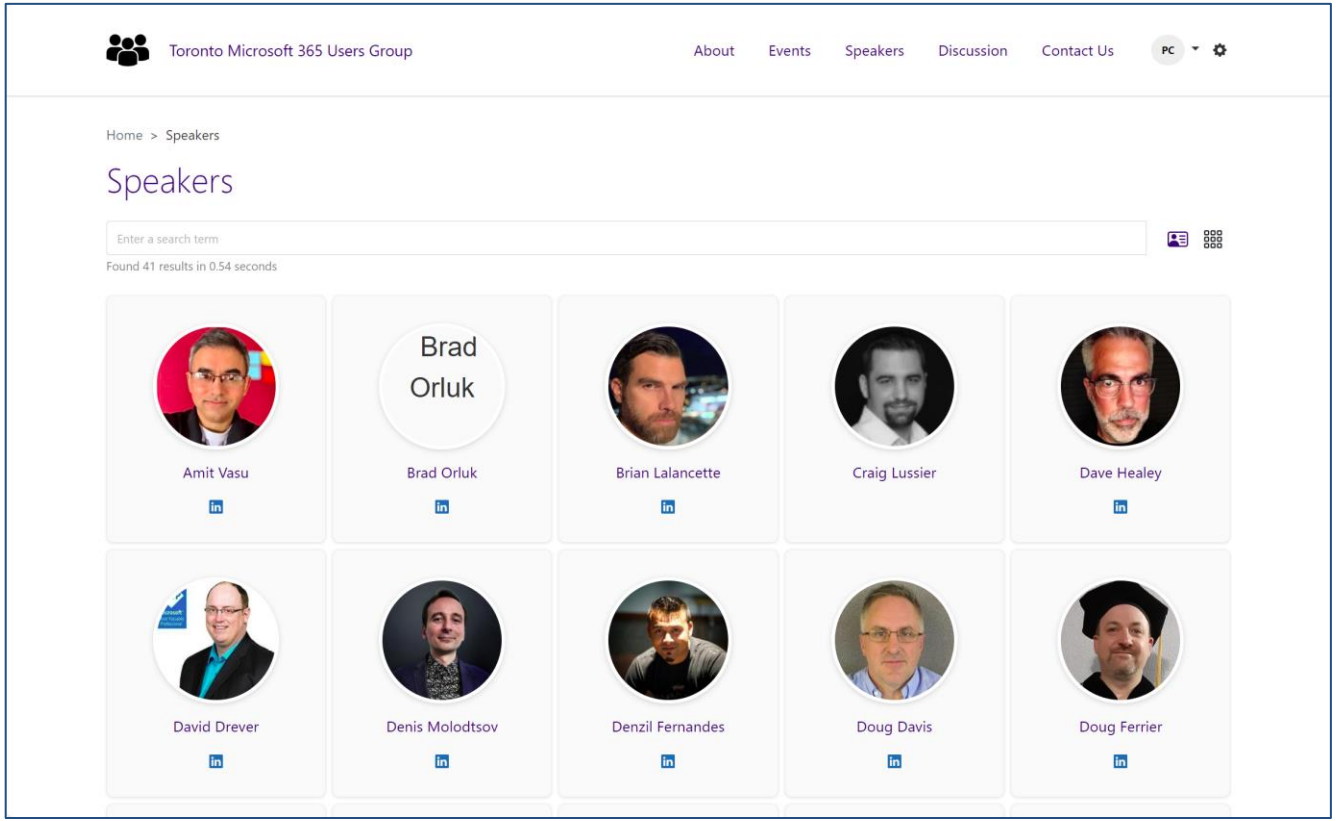

ExtranetUserManager
Platform
Solutions
Resources
Pricing
About
Support
Test Ride →
PC
Home > About Us > Staff Directory
Staff Directory
Enter a search term
Found 25 results in 0.52 seconds
Aaron Silver
Developer
Alina Raza
UI / UX Developer
Amanda Da Costa
Executive Assistant
Amanda Diflorio
Digital Marketing Specialist
Anish Birju
Developer
Artur Pinheiro
Developer
Danish Raza
Business Analyst
Denesh Sohan
Director of Engineering
Devraj Banerjee
Customer Success Specialist
Don Harkness
Chief Financial Officer
Eman Abdeen
Quality Assurance and Support Development
Faye Yuan
Senior Accountant
Gurinderpal Dhillon
Senior Developer
Irina Kenic
Marketing Specialist
Jeff Clement
Creative Director
Joel Stickland
Sales Development Representative
Julian Mills
VP Product and Marketing
Kate Carson
Developer
Logan Guest
VP Sales and Customer Success
Maha Elkoshairi
Senior Consultant



https://mbncanada.ca



https://www.citywindsor.ca



https://tm365.org

Key Points

- Search is a key component of a chat experience
- Think about grounding and security
- There are many ways to do this within Microsoft
  - Copilot
  - Copilot Studio
  - Custom dev and Azure
- Azure AI Search is incredibly powerful

Resources

- AI Community Conference - Toronto 2024 | Extranet User Manager
- Azure AI Search and Open AI Webinar | Extranet User Manager
- Get started with the chat using your own data sample for JavaScript - JavaScript on Azure | Microsoft Learn
- Azure AI Search documentation | Microsoft Learn
- Simple query syntax - Azure AI Search | Microsoft Learn
- Pricing - Azure AI Search | Microsoft Azure
- Azure OpenAI Service - Pricing | Microsoft Azure

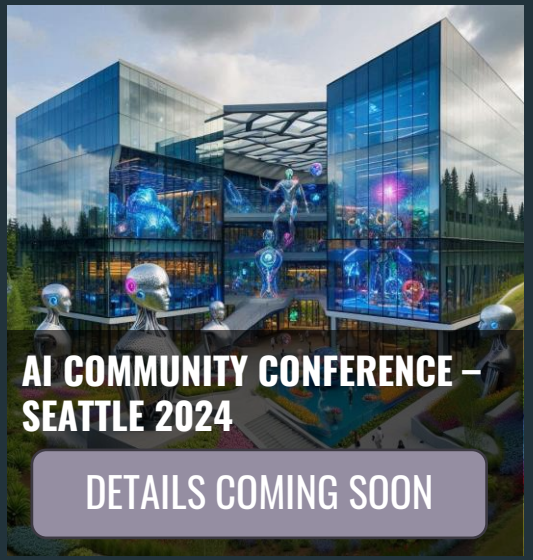
Speaker Evaluations

Please complete an evaluation for each session you are attending. This will help us in providing our speakers with valuable feedback. You will also increase your chances of winning a raffle prize

Upcoming Events

Webinar: Teams and Microsoft 365 Sharing is Caring
When: July 18th, 10-10:45AM EST
What: Microsoft 365 sharing scenarios, controlling access and rights with Entra ID and Microsoft Purview, leveraging custom and third-party options via the Graph API, and highlights use cases like learning portals, virtual data rooms, committees, meetings, dashboards, and reports.

Register now!

Thank you for attending

We hope to see you at one of our upcoming events